経済学研究年報（早大）第27号　1988年2月

共通主成分分析の特性と応用

勝　浦　正　樹

はじめに

共通主成分分析（Common Principal Component Analysis；CPCA）とは，k 個のグループに対して一括して主成分分析（Principal Component Analysis；PCA）を行う方法であり，主成分分析を一般化したものであるといえる。すなわち，1個のグループ（$k=1$）に対する共通主成分分析が，通常の主成分分析にあたる。概念的には Krzanowski [13] において，グループ間の主成分分析について議論されているが，k 個のグループに対して，推定方法その他を統計学的に定式化したのは，Flury である（Flury [4]）。

そこで，以下では，まず共通主成分分析について概説した上で，その意義や問題点を論ずる。そして応用例として，日本の府県別データを用いて，産業構造の変化の実証的な分析を行う。とりわけ，府県別データというクロスセクションデータの時系列的な変化をとらえる方法として，共通主成分分析が有効であり，時系列データとクロスセクションデータの結合（pooling）の1つの技法として，共通主成分分析が利用できることを示す。

1　共通主成分分析概説

(1)　共通主成分分析の目的

いま異なった k 個のグループが存在し，それぞれについて p 個の同一の変数が測定されている状態を考えよう。例えば，ある学校のいくつかのクラス（k 個）に対して，各生徒の国語・数学・英語・理科・社会という5教科（$p=5$）

の得点が与えられているような状態である。ここで，各クラス内の人数は異なってもかまわないが，対象となる変数（教科）は共通とする。各生徒の単純な合計点ではなく，各クラスごとに主成分分析を用いて，教科ごとのウェイトを決めたり，主成分得点を計算したりすることが可能であり，有効である。しかし主成分分析の結果はクラスごとに異なっているので，共通の主成分得点を計算するためには，クラスの枠を外して，全員に対して主成分分析を行わなければならない。そこで，クラスというグループの構造を残したままで主成分分析を行うのが，共通主成分分析である。つまり共通主成分分析というのは，グループが存在したときに，それぞれのグループに共通の構造を考え，各グループに対して同時に主成分分析を適用する方法であり，グループ間の類似性や相違を明らかにし，共通の測度での比較を可能にする。

(2) 主成分分析と共通主成分分析の相違

主成分分析と共通主成分分析はこのような関係にあるが，k 個のグループに対して個別に主成分分析を行うのと，同時に共通主成分分析を行うのでは，どんな違いがあるのだろうか。$k=1$ であれば，もちろん両者は一致する。また，k 個のグループに対する個別の主成分分析の結果と，共通主成分分析の結果の比較は重要である。ここではとりあえず，形式的な（数学上の）定式化の違いを示すことにより，両者の相違をはっきりさせる。

通常の主成分分析では，n 個の観察値に対し，p 個の変数についてのデータが与えられている状態を考える。そして，その分散・共分散行列 $\boldsymbol{\Sigma}(p\times p)$ に対して，

$\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B}=\boldsymbol{\Lambda}$，但し，$\boldsymbol{\Lambda}$ は対角行列

という対角化を行うような，直交行列 $\boldsymbol{B}$ を見つける固有値問題に帰着する。もちろん，$\boldsymbol{\Lambda}$ の対角要素は $\boldsymbol{\Sigma}$ の固有値であり，$\boldsymbol{B}$ は，その固有値に対応する固有ベクトルを列ベクトルにもつ行列である。これに対して k 個のグループ（観察数は n_1，……，n_k とする）が存在する場合，分散・共分散行列も k 個存在する。そこで共通主成分分析は，k 個の分散・共分散行列 $\boldsymbol{\Sigma}_1$，……，$\boldsymbol{\Sigma}_k$

を，同時に対角化する直交行列 $\boldsymbol{B}$ を考える。すなわち，

$$\boldsymbol{B}'\boldsymbol{\Sigma}_1\boldsymbol{B}=\boldsymbol{\Lambda}_1,\ \boldsymbol{B}'\boldsymbol{\Sigma}_2\boldsymbol{B}=\boldsymbol{\Lambda}_2,\ \cdots\cdots,\ \boldsymbol{B}'\boldsymbol{\Sigma}_k\boldsymbol{B}=\boldsymbol{\Lambda}_k$$

と共通主成分分析は定式化される。ここで，共通の $\boldsymbol{B}$ という行列で同時対角化を行うのであって，対角化された行列 $\boldsymbol{\Lambda}_i$ は共通ではない，ということは重要である。また，$\boldsymbol{\Sigma}_1\cdots\cdots$，$\boldsymbol{\Sigma}_k$ が特定の条件を満たす場合を除いて，一般に $\boldsymbol{\Lambda}_1, \cdots\cdots, \boldsymbol{\Lambda}_k$ は厳密に対角行列にならない。ある１つの行列を対角化する直交行列の存在は証明されるが，複数の行列を同時に対角化することは，数学的に証明されない。そこで共通主成分分析では，ある基準を設けて，$\boldsymbol{\Lambda}_1$，$\cdots\cdots$，$\boldsymbol{\Lambda}_k$ ができる限り対角化されるように $\boldsymbol{B}$ を推定する。つまり近似的な対角化であって，対角行列に近づくほど各グループに共通の構造があるといえる。また共通の構造があれば，$\boldsymbol{B}$ によってグループの共通性が見い出せるのである。

(3) 共通主成分分析における推定方法

Flury が示した共通主成分分析の推定方法(1) をまとめておくと，以下のようになる。まず $X_i \sim N_p(\boldsymbol{\mu}_i,\ \boldsymbol{\Sigma}_i)$，$i=1,\ \cdots\cdots,\ k$ とする。ここで，X_i は $p\times n_i$ の観察値行列で，独立に分布すると仮定する。S_i を標本不偏共分散行列として，$\boldsymbol{\Sigma}_1$，$\cdots\cdots$，$\boldsymbol{\Sigma}_k$ の最尤推定量を求めてみる。具体的には尤度関数，

$$L(\boldsymbol{\Sigma}_1, \cdots\cdots, \boldsymbol{\Sigma}_k)=C\times\prod_{i=1}^{k}\exp\left[tr\left(-\frac{n_i}{2}\boldsymbol{\Sigma}_i{}^{-1}S_i\right)\right]|\boldsymbol{\Sigma}_i|^{-\frac{n_i}{2}}$$

但し，C は $\boldsymbol{\Sigma}_i$ に依存しない定数(2)

を最大化するかわりに，

$$g(\boldsymbol{\Sigma}_1, \cdots\cdots, \boldsymbol{\Sigma}_k)=-2\log L(\boldsymbol{\Sigma}_1, \cdots\cdots, \boldsymbol{\Sigma}_k)+2\log C$$

を最小化することにより，$\boldsymbol{\Sigma}_1, \cdots\cdots, \boldsymbol{\Sigma}_k$ の最尤推定量を求める。ここで，共通主成分仮説 (hypothesis of common principal component's)

$$H_c:\ \boldsymbol{B}'\boldsymbol{\Sigma}_i\boldsymbol{B}=\boldsymbol{\Lambda}_i\ (\text{対角行列}),\ i=1, \cdots\cdots, k$$

という仮説をたてる ($\boldsymbol{B}$ は $p\times p$ の直交行列)。そして H_c 仮説が成立するという条件のもとで $g(\boldsymbol{\Sigma}_1, \cdots\cdots, \boldsymbol{\Sigma}_k)$ の最小化を考える。過程は省略し，(3) 結果だけを示すと

$$\boldsymbol{\beta}_l'\left(\sum_{i=1}^{k} n_i \frac{\lambda_{il}-\lambda_{ij}}{\lambda_{il}\lambda_{ij}} \boldsymbol{S}_i\right)\boldsymbol{\beta}_j=0,\ \text{但し}\ l,\ j=1,\cdots\cdots,p,\ l\neq j$$

という $p(p-1)/2$ 個の方程式を解けばよい。ここで，$\boldsymbol{\beta}_l$ は $\boldsymbol{B}$ の第 l 列ベクトル，λ_{il} は $\boldsymbol{\Lambda}_i$ の第 l 対角要素である。もちろん $\boldsymbol{B}'\boldsymbol{B}=\boldsymbol{I}_p$ という制約条件のもとで，方程式は計算される。この方程式によって推定された $\boldsymbol{\beta}, \lambda$ にへをつけて表わせば，

$$\hat{\boldsymbol{B}}=(\hat{\boldsymbol{\beta}}_1,\cdots\cdots,\hat{\boldsymbol{\beta}}_p),\ \ \hat{\boldsymbol{\Lambda}}_i=\mathrm{diag}\{\hat{\lambda}_{ij}\},\ \ \hat{\boldsymbol{\Sigma}}_i=\hat{\boldsymbol{B}}'\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{B}}$$

と最尤推定量を表わすことができる。

また，H_c 仮説の検定には，尤度比検定を用いることができる。検定統計量には，

$$\chi^2=-2\log\frac{L(\hat{\boldsymbol{\Sigma}}_1,\cdots\cdots,\hat{\boldsymbol{\Sigma}}_k)}{L(\boldsymbol{S}_1,\cdots\cdots,\boldsymbol{S}_k)}$$

を用いて，$\min n_i\to\infty$ ならば，χ^2 が自由度 $(k-1)p(p-1)/2$ の χ^2 分布する[4] ことを利用して，検定を行えばよい。

さてここで問題となるのは，共通主成分分析がどの程度うまく行くかである。なぜならば，H_c 仮説は厳密には成立しないのが一般的だからである。H_c 仮説は，$\boldsymbol{\Lambda}_i$ が完全に対角化されれば成立する。そこで Flury は，その対角行列からの乖離の測度として，

$$\phi(\boldsymbol{F}_i)=|\mathrm{diag}\,\boldsymbol{F}_i|\,/\,|\boldsymbol{F}_i|$$

という基準を考えている。[5] 但し，$\boldsymbol{F}_i=\hat{\boldsymbol{B}}'\boldsymbol{S}_i\hat{\boldsymbol{B}},\ \hat{\boldsymbol{\Lambda}}_i=\mathrm{diag}(\boldsymbol{F}_i)$ である。この基準は $\boldsymbol{F}_i$ が対角行列ならば1になり，$\boldsymbol{F}_i$ が対角行列から離れるほど大きくなる。$\boldsymbol{F}_i$ を k 個について考えた基準は，

$$\Phi(\boldsymbol{F}_1,\cdots\cdots,\boldsymbol{F}_k\,;\,n_1,\cdots\cdots,n_k)=\prod_{i=1}^{k}[\phi(\boldsymbol{F}_i)]^{n_i}$$

となり，この Φ を最小にする $\boldsymbol{F}_i=\hat{\boldsymbol{B}}'\boldsymbol{S}_i\hat{\boldsymbol{B}}$ と，上で述べた最尤推定量は一致する。[6] 逆に，ϕ という基準をもとに，共通主成分分析は出発しているとも言えよう。

(4) 共通主成分分析の有効性

共通主成分分析が有効となる状況は，H_c 仮説が成立している状態，またはそれに近い状態であることは明らかである。ではもとのデータがどのような状態にあるときに，H_c 仮説は成り立つのだろうか。結論から示せば，グループごとの分散・共分散行列が比例的である場合である。すなわち，$\boldsymbol{\Sigma}_i=c_i\boldsymbol{\Sigma}_0$ が成立している場合に，$\boldsymbol{\Lambda}_i$ は完全に対角化されることが証明される[7]（右辺にある $\boldsymbol{\Sigma}_0$ を他の分散・共分散行列にしても一般性は失わない)。つまり，H_c 仮説と比例性の仮定は同値である。2個，またはそれ以上の分散・共分散行列が等しいという仮説の検定には，いくつかの方法がある[8]が，それらが等しいという仮説は，R. A. Fisher の Iris の例に代表されるようにほとんど棄却されてしまう。[9] $\boldsymbol{\Sigma}_i=c_i\boldsymbol{\Sigma}_0$ という比例性に関する検定は，まだ十分に議論されているとはいえないが，$\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma}_0$ よりは緩い仮定になるだろう。また，χ^2 検定によって H_c 仮説が棄却されるような場合には，分散・共分散行列の比例性の仮説が棄却されるといってもよいだろう。つまり共通主成分分析は，分散・共分散行列の比例性の検定にも利用可能である。

では比例的な分散・共分散行列とは，どんな意味をもつのだろうか。あるグループのデータが，他のグループのデータの一定倍になれば，もちろんそれが当てはまっているが，そのような状態は，実際にはほとんど存在しないだろう。ところが共分散が一定倍であっても，相関係数は変わらないはずである。つまり比例性の仮定というのは，共通の相関構造をもっているということに対応する。従って共通主成分分析が有効となる状況は，グループが存在し，変数の観測値が異なっていても，変数間の相関構造がグループごとに一定，または類似している状況であるといえよう。もちろん，その相関構造の同質性については，先に述べた χ^2 検定などによって検証が可能である。正確に相関係数が一致していなくても，H_c 仮説が棄却されないかぎり，また共通した相関構造をもっているということを先験的に仮定している場合には，共通主成分分析は有効でありうる。

(5) 共通主成分分析の問題点

以上のように共通主成分分析をまとめることができるが，その問題点も多い。特に，主成分分析に比べてデータの制約条件がきびしくなる。まず，データに多変量正規分布を仮定している。これは生物学的なデータならともかく，経済データにはまずあてはまらないだろう。この仮定は，最尤推定量を導くためのものであるから，推定方法を変えれば解決できよう。またそれと関連して，対角化の基準が ϕ という基準であることも問題である。一般的には，対角化の基準として，非対角要素が０にできるだけ近いことが考えられる。例えば，非対角要素の２乗和を考え，それが最小となるような最小２乗推定量を考えることが可能であろう。(10) その場合，正規性を仮定する必要がなくなる可能性があり，より実践的になるだろう。またロバストな推定方法として，共通主成分分析を適用する分散・共分散行列に，M推定量などを用いる方法も考えられる。(11)

その他，主成分分析と比較した問題点としては，あくまでも近似的な対角化であること，数値計算上の問題，(12) 得られた（共通）主成分が完全に独立ではないことなどがあげられる。さらに，通常の主成分分析でも存在する問題だが，得られた主成分の解釈，次元の減少の可能性などといった分析上の問題もある。(13)

このような問題点を改良しながら，共通主成分分析をより応用範囲の広いものにしていくことが期待される。

注(1) Flury [4].
(2) Anderson [1] pp. 245-251 などを参照のこと。
(3) Flury [4] pp. 892-893.
(4) Rao [16] 邦訳378-382ページ。
(5) Flury and Gautschi [9] p. 169.
(6) Flury and Gautschi [9] pp. 169-171.
(7) Flury [6] pp. 29-31.
(8) Anderson [1] chap. 10に，分散・共分散行列の検定の問題が論じられている。
(9) Anderson [1] p. 446, Flury [4] pp. 894-895.

(10) $\phi(\boldsymbol{F}_i)=\sum_{\substack{l=1\\ l\neq j}}^{p}\sum_{j=1}^{p} f_{ij}^2$

$$\Psi(\boldsymbol{F}_1,\cdots\cdots,\boldsymbol{F}_k;n_1,\cdots\cdots,n_k)=\sum_{i=1}^{k} n_i[\phi(\boldsymbol{F}_i)]$$

として，Ψ の最小化を考えればよい。

(11) 主成分分析におけるロバストな推定については，Devlin et al. [3] で実験的に検証されている。

(12) 共通主成分分析における数値計算上の問題（収束可能性，解の一意性など）については，Flury and Gautschi [9] pp. 171–184 を参照のこと。

(13) 共通主成分分析における次元の減少の問題については，Flury [7] pp. 59–63 を参照のこと。その他，共通主成分分析については，Flury の一連の論文や Krzanowski [14] で，様々な問題が議論されている。

2. 共通主成分分析による産業構造の変化の分析

では，共通主成分分析を実際のデータに適用すると，どうなるだろうか。通常の主成分分析が産業構造の変化をとらえるのに適切な方法である，ということをさらに発展させて，その時系列的な変化を追うために共通主成分分析を応用してみる。

(1) 主成分分析による産業構造の変化の分析

産業構造の変化をとらえる古典的な方法としては，ペティ・クラークの法則がある。これは，経済が発展する（具体的には所得水準が上昇する）に従って，経済における比重・重要性が第1次産業から第2次産業，さらには第3次産業へと移行するという経験的な法則である。ここではペティ・クラークの法則を検証することが目的ではないが，所得水準と産業構造の関係を分析することは有意義であろう。産業構造を変化させる1つの要因として，所得水準が考えられるからである。両者の関係の分析方法としては，所得変数を独立変数，各産業の就業構成比を従属変数にした回帰分析が考えられる（以下では議論を単純にするため，産業構造を，就業者の構成比でとらえるものとする(14)）。回帰分析の手法を用いるとき，産業を何らかの基準で q 個に分類したとすれば，例えば，

$$\frac{N_i}{N}=a+b\frac{Y}{N}$$

但し，N_i：第 i 産業の就業人口（$i=1, \cdots\cdots, q$）

N：全就業人口

Y：国民または県民所得

という式を用いて推定を行うことになる。もちろん，従属変数には，各産業の就業構成比だけでなく，それらの合成変数[15]を用いても良いが，問題点は，各産業の変数を同時に扱うことが困難なことである。すなわち，$N_1/N, \cdots\cdots, N_q/N$, Y/N を同時に回帰式に取りこむことが，多重共線性の問題などから技術的にも難しい。そこで主成分分析を用いれば，それらの変数を同時に扱い，各変数の相関関係（とりわけ所得変数と就業構成比の関係）を相互依存的に知ることができる。こうしたことから，産業構造の分析に主成分分析を用いることが有効だといえよう。

ここで主成分分析に用いるデータは，クロスセクションデータか時系列のデータのいずれかに限定される。ところが，各時点のクロスセクションデータを1つのグループと見なし，k 個の時点に対して共通主成分分析を行えば，クロスセクション的な相関構造と時系列的な相関構造を同時に考慮できる。つまり共通主成分分析によって，ある種の pooling を行うことになる。もちろん共通主成分分析を用いるには，いくつかの前提条件が必要となる。正規性の仮定は別にして，データの状態は，個別の主成分分析と比較し，妥当性をみればよいだろう。特に，固有値・固有ベクトルを主成分の解釈などと関連させた上で比較し，共通主成分分析と個別の主成分分析で，かけ離れた結果が出てくる状態は好ましくない。また，グループ間（時点間）で共通の相関構造をもっているかどうかということは，産業構造と所得水準の間に何らかの系統的な関係（または経済理論）が，対象期間中には存在すると仮定すればよいし，共通主成分分析は，そうした関係が存在するかどうかを確かめるのに役立つことになる。

(2) 結果と考察

分析の対象は，日本の都道府県別のデータである。[16] 所得変数には1人あたり県民所得（全国平均=100とした指数，経済企画庁『県民経済計算年報』による），産業別の就業構成比には農林業，製造業，卸売・小売業，サービス業のデータ（総務庁統計局『就業構造基本調査』による）を用いた。クラーク流の第1次・第2次・第3次産業という分類ではやや大雑把すぎるので，もう少し細かく分析するために代表的な4つの産業に分けている。もちろん，第3次産業に含まれる各産業は，性質がかなり異なり，職種も複雑化・多様化の傾向が見られるので，この分類で産業構造を把握しきれるとは言い難い。しかしここでは，共通主成分分析の適用可能性を示すことに重点を置いているので，ある程度伝統的な方法に沿って，分析を進めていくことにする。対象年度は昭和46年，52年，57年である。[17] 2度のオイルショックを間にはさんでも，就業構造と所得の関係が安定的に保たれているかどうか，各都道府県がどのような変化をしているのかを考察する。データ数は昭和46年は，40,[18] 52年・57年は46（沖縄県を除く）である。

表1は，各年の主成分分析の結果を示している。[19] 各年とも第2主成分までの累積寄与率が90%近くになっているので，この2つの主成分についてみてみ

表1　昭和46・52・57年の主成分分析の結果

	第1主成分			第2主成分		
	昭和46年	昭和52年	昭和57年	昭和46年	昭和52年	昭和57年
X_1	0.5524	0.5570	0.5740	0.0166	0.0400	0.0916
X_2	−0.4396	−0.3916	−0.3176	−0.4874	−0.5849	−0.6327
X_3	−0.4463	−0.4696	−0.4691	0.4008	0.3008	0.3736
X_4	−0.1196	−0.1757	−0.2262	0.7705	0.7511	0.6595
X_5	−0.5368	−0.5339	−0.5463	−0.0887	−0.0411	−0.1297
固有値	3.1500	2.9987	2.7248	1.4496	1.3713	1.6854
寄与率	0.6300	0.5998	0.5450	0.2899	0.2743	0.3371

X_1：農林業就業構成比　X_2：製造業就業構成比　X_3：卸売・小売業就業構成比　X_4：サービス業就業構成比　X_5：1人あたり県民所得

表2　共通主成分分析の結果

	第1成分	第2成分
X_1	0.5585	0.0401
X_2	−0.3991	−0.5646
X_3	−0.4584	0.3799
X_4	−0.1644	0.7267
X_5	−0.5400	−0.0850

固有値と寄与率（ ）

	第　1	第　2
昭和46年	3.1429 (0.63)	1.4485 (0.29)
昭和52年	2.9981 (0.60)	1.3619 (0.27)
昭和57年	2.7105 (0.54)	1.6873 (0.33)

よう。まず第一に，各固有ベクトルの要素の大きさの変動は小さく，各変数の相関関係が安定していることが分かる。また主成分を解釈すると，第1主成分は農業化を，第2主成分はサービス化を示しているといえよう。第1主成分をみると，X_1 の値が大きく，X_2, X_3, X_5 と符号が逆になり，負の相関をもっているのが分かる。特に X_1 と X_5 の負の相関が強い。つまり所得が上昇すると，農林業の就業構成比は下がり，製造業や卸売・小売業の就業構成比が上がることを意味している。また第2主成分では，サービス業や卸売・小売業のウェイトが高くなっており，製造業とは負の相関を持っていることが分かる。所得とサービス化はこれからみるとそれほど強い関係になく，むしろ非工業化がサービス化をもたらしているようである。つまり所得の上昇が第1次産業から第2次産業，卸売・小売業への移行をもたらし，第2次産業が飽和状態になれば，サービス業へ移行していくという図式が，ある程度確認できる。

次に，共通主成分分析の結果を**表2**に示す。[20] 時間を通じても共通の相関構造があることが，個別の主成分分析でも確認されているので，結果は良好である。表1の3つの主成分分析を，うまくまとめた結果になっている。従って，各主成分の解釈も同様にできる。一応 χ^2 を計算してみると，$\chi^2=18.27$ となり，$\chi^2_{0.05}(20)=31.41$（自由度は，$k=3$, $p=5$ より $(k-1)p(p-1)/2=20$ である）なので，H_c 仮説は棄却されない。

各変数の状態をもう少し分かりやすくするために，因子負荷量を第1主成分と第2主成分の座標上へプロットしてみよう（**図1**）。[21] この図から，農林業と

図1　因子負荷量のプロット

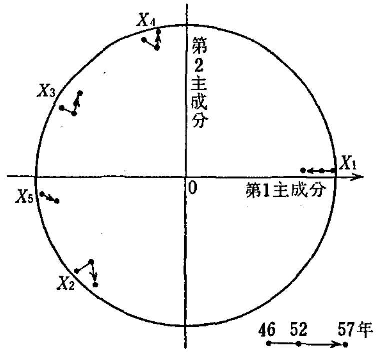
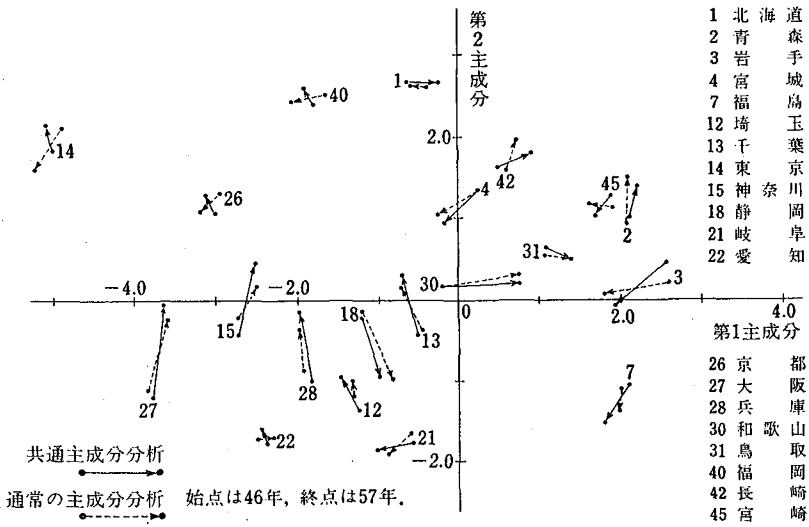


図2　主成分得点の第1・第2主成分へのプロット(22)



所得は全く逆の関係にあることが分かる。また，所得は製造業や卸売・小売業への波及が強く，サービス業への影響が小さいことが読みとれる。さらに因子負荷量は，年度によって安定しており，相関構造に共通するものがあることの裏付けとなる。

では，各都道府県の主成分得点を計算してみよう。**図2**は，代表的な都道府

県についての，主成分分析による主成分得点（破線）と，共通主成分分析による共通主成分得点（実線）の昭和46年から57年までの変化を描いている。[23] ここで注意したいのは，個別の主成分分析を異なった年度で比較することは誤った方法であることである。標準化した変数を用いても，異なった時点では各変数に対するウェイトは違っているので，厳密な比較はできないはずである。[24] 従って，グループ間の比較（ここでは時系列の比較）の条件を整えるためには，共通主成分分析による共通のウェイトが必要となる。

第１・第２主成分の解釈から，図の上に行くほどサービス化が進んでおり，下に行くほど工業化が進んでいる。また左に行くほど所得が高く，右に行くほど農業が中心である。従って，主成分得点の変化の方向が左上に行くほど，第３次産業化による所得の増加，左下に行くほど工業化による所得の増加が進んでいることになる。通常の主成分分析によれば，東京（14）・京都（26）・福岡（40）などはサービス化は進んでないことになる。これは昭和46年より57年においては所得のウェイトが高くなっているので，サービス化よりも所得増による影響が強く出てしまったためである思われる。このような状況も，共通主成分分析を用いれば，各年を共通の尺度で比較できるので解消されている。この図からサービス化が進んでいる県（埼玉・千葉・東京・神奈川・京都・大阪・兵庫・福岡など），工業化による所得増が見られる，あるいはサービス化が十分に進んでいない県（宮城・福島・静岡など），農業の比重が高い県[25]（青森・和歌山・鳥取・宮崎など）というように分類できるだろう。これらの結果を，もとのデータによって確かめてみよう。宮城，福島，静岡の製造業の就業構成比は，昭和46年から57年にかけてそれぞれ，14.6→16.6%，20.5→23.0%，31.7→31.7%であり，全国平均が27.2→24.6%と減少していることを考慮すれば，妥当な結果と思われる。静岡については，昭和57年のサービス業の就業構成比が18.0%で，全国平均の19.3%を下回ったことも影響している（これらの県の製造業の就業人口の伸びは，全国平均を大きく上回っており，サービス業の伸びは下回っていることからも裏付けられる)。また農業の比重が高い県についても，同様のことが言える。またこれらの県では，所得の相対的位置が46年か

ら57年にかけて下がっていることも，図２での右方向への動きに影響している（例えば和歌山は，所得は91.5から81.5に下がっている)。このように，それぞれの結果は実情に一致しているといえ，共通主成分分析による産業構造の分析が有効であることを示していると思われる。

このように，共通主成分分析を用いることにより，クロスセクション的な相関構造と時系列的な相関構造を同時に扱い，最近の日本経済における産業構造と所得の変化の関係について，１つの傾向を見い出すことができた。特に，サービス化が必ずしも所得水準と密接に結びついていない，という結果は特徴的である。しかし，ここで行った分析にもいくつか問題がある。まずデータが正規分布をしてないことがあげられる。また，より厳密な産業構造の分析のためには，ここで取り上げた以外の変数を考慮する必要があるだろう。(20) さらに，主成分分析においては，特定の変数の因果の方向は特定化されていないので，所得が産業構造の変動要因になっているかどうかは，主成分分析からは何もいえない。回帰分析のように，特定のモデルを設定していないからである。その他にも，産業の分類のしかた，他の産業の構成比を用いたらどうなるかといった問題もある。とはいえ，産業構造をこのような方法で分析することが可能であるということは，十分示せただろう。

注(14) 産業構造をとらえるには，この他に生産額の構成比などが一般的である。

(15) 例えば $(N_1+N_2)/N_3$ といったように合成すればよい。しかし各産業どうしの相関をとらえることはできない。

(16) 篠原［17］第７章で，府県別産業構造の1970年代前半までの分析が行われている。

(17) この３年間は，オイルショック前後を含む期間のうち，府県別就業者が公表される就業構造基本調査が行われた年のなかから選んだ。なお，12年間だけでは，変動が比較的小さい産業構造の分析には，不適切であるかもしれない。しかし，クロスセクションデータと時系列データの両方を用いることにより，その問題はある程度解決されるだろう。

(18) 新潟，茨城，山梨，三重，島根，愛媛，沖縄の各県を除く。これは『県民経済計算年報』で，昭和47年までこれらの県のデータが与えられていないことによる。

(19) 主成分分析ならびに共通主成分分析は相関行列に対して行った。

(20) 共通主成分分析の計算のためのプログラムは，Flury and Constantine [8] を

参照した。なお，共通主成分分析の結果で，通常の主成分分析と同様に，主成分，固有値，寄与率などという用語を用いているが，厳密には主成分分析におけるそれらとは異なった性質を持つもので，正しい用語法ではない。

(21) 因子負荷量というのは，主成分と各変数との相関係数である。図1は，共通主成分分析の結果を用いて，因子負荷量を算出した。

(22) ここで用いている都道府県につけられた番号は，『県民経済計算年報』に基づいており，通常の順序とは多少異なっている。

(23) 図の繁雑さを避けるために，昭和52年度は省略し，変化を1つの線分で表わした。52年のデータを記入しても，その線分上にくるとは限らない。しかし，だいたいその近傍にプロットされ，変化の方向に大きな変動はない。

(24) 例えば，昭和46年の X_1 のウェイトは0.5524，52年だと0.5740となっている。その数値を，そのままもとのデータに乗じて主成分得点を計算しても，得られた得点のもつ意味は異なってくる。主成分分析を行うときに，陥りやすい誤りである。

(25) もちろん，他の県と比較して相対的に農業の比重が高いという意味である。

(26) 例えば，人口構造を表わす変数や，生産性を表わす変数などが考えられる。

おわりに

以上，共通主成分分析の概要とその適用可能性について述べてきたが，実証分析に利用するためには，前提条件を吟味し，個別の主成分分析との比較を行うことが重要になる。また，推定方法を改良することによって，前提条件を緩めることも可能であろう。ここでは産業構造の分析に利用したが，その他にも，景気分析や計量経済学における pooling の技法(27) としてなど，いろいろな適用範囲が考えられる。景気分析に主成分分析を用いる(28) 場合には，対象期間を景気の1循環ごとに分け，それぞれを1グループと見なして共通主成分分析を適用すれば，複雑な個別指標間の相関構造がよりはっきりと出てくるだろう。

しかし共通主成分分析という方法はまだ一般的ではなく，統計理論・実証分析の両方の立場から，いろいろな角度で議論していく必要があるだろう。単なる主成分分析の一般化というだけでなく，分散・共分散行列の検定，回帰分析への適用など様々なトピックを含んだ問題であるといえよう。

注(27) pooling technique については，Balestra and Nerlove [2]，Johnston [11] pp. 396-407, Maddala [15] などを参照のこと。

(28) 景気分析を含め時系列データに主成分分析を適用することの理論的な裏付けは，刈屋 [12] の MTV モデルによって与えられている。

参考文献

[1] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, 1984, 675pp.

[2] Balestra, P. and Nerlove, M., "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica*, vol. 34, 1966, pp. 585–612.

[3] Devlin, S. J., Gnanadesikan, R. and Kettenberg, J. R., "Robust Estimation of Dispersion Matrices and Principal Components", *Journal of the American Statistical Association*, vol. 76, 1981, pp. 354–362.

[4] Flury, B. N., "Common Principal Components in k Groups", *Journal of the American Statistical Association*, vol. 79, 1984, pp. 892–898.

[5] Flury, B. N., "Asymptotic Theory for Common Principal Component Analysis", *Annals of Statistics*, vol. 14, 1986, pp. 418–430.

[6] Flury, B. N., "Proportionality of k Covariance Matrices", *Statistics and Probability Letters*, vol. 4, 1986, pp. 29–33.

[7] Flury, B. N., "Two Generalizations of the Common Principal Component Model" *Biometrika*, vol. 74, 1987, pp. 55–69.

[8] Flury, B. N. and Constantine, G., "Algorithm AS 211: The FG Diagonalization Algorithm", *Applied Statistics*, vol. 34, 1985, pp. 178–183.

[9] Flury, B. N. and Gautschi, W., "An Algorithm for Simultaneous Orthogonal Transformation of Several Positive Definite Symmetric Matrices to Nearly Diagonal Form", *SIAM Journal of Scientific and Statistical Computing*, vol. 7, 1984, pp. 169–184.

[10] Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology*, vol. 24, 1933, pp. 417–441 and 498–520.

[11] Johnston, J., *Econometric Methods*, 3rd ed., McGraw-Hill, 1984, 563pp.

[12] 刈屋武昭『計量経済分析の考え方と実際』 東洋経済新報社 1986年, 189ページ。

[13] Krzanowski, W. J., "Between-Group Comparison of Principal Components", *Journal of the American Statistical Association*, vol. 74, 1979, pp. 703–707.

[14] Krzanowski, W. J., "Principal Component Analysis in the Presence of Group Structure", *Applied Statistics*, vol. 33, 1984, pp. 164–168.

[15] Maddala, G. S., "The Use of Variance Components Models in Pooling Cross-Section and Time-Series Data", *Econometrica*, vol. 39, 1971, pp. 341–358.

[16] Rao, C. R., *Linear Statistical Inference and Its Applications*, 2nd ed., John

Wiley & Sons, 1973, 625pp（奥野忠一他訳 『統計的推測とその応用』 東京図書 1977年，586ページ).

[17] 篠原三代平 『産業構造論』 筑摩書房 1976年，409ページ。

[18] 総理府統計局 『人口の就業状態と産業構成』 昭和55年国勢調査モノグラフシリーズ No. 4 日本統計協会 1983年，218ページ。

[19] 田中豊・脇本和昌 『多変量解析法』 現代数学社 1983年，296ページ。

1987. 9. 30 提出

（博士後期課程第2年度生）